Jour. Agric. Res. China (中華農業研究) 3(4) : 475~483 (1985)

Component Analysis of Heterozygous Diallel Crosses[1]

Ching Liu[2]

Abstract : By constructing separate models for selfings and crossbreds based on their genotypic values, the genetic components analysis of Hayman and the combining ability analysis of Griffing for a diallel crosses were proved to be essentially equivalent and were combined to provide estimates of various genetic parameters of the sampled population from a heterozygous diallel. Difference in estimation of genetic parameters between homozygous diallel and heterozynous diallel were indicated.

Introduction

Any set of matings among p individuals, strains or inbred lines which can be arranged into sets of symmetrical two-way classifications by parentage may be called a diallel crosses. Many and seemingly different diallel analyses have been presented in the literature. Diallel analyses differ in four main ways-in the sampling assumptions of the experimental materials, in the postulated underlying genetic mechanism, in the methods of estimation, and in the parental heterozygosity.

Yates (1947) has given a statistical analysis by using the principles of analysis of variance for the factorial experiment; this has been modified and stated in terms of biometrical genetic parameters by Jinks and Hayman (1953) and further by Hayman (1954a, b, 1958, 1960) and Jinks (1954). Hayman (1954a) has defined five components (h, D, H_1, H_2, and F) of genetic variance in a homozygous diallel crossing system exhibiting additive and dominance, but no epistatic variance. The method of Jinks and Hayman (1953) emphasizes the analysis of genetic variances for a diallel crosses of inbred lines. They used diagrammatic methods to estimate the ratio of additive genetic variance to non-additive genetic variance in order to determine the mean level of dominance for a quantitative character. Dickinson and Jinks (1956) have extended the method of Jinks and Hayman to diallel crosses involving parental heterozygosity but the use of their method requires information about the parents which is not generally available.

Kempthorne (1956) has discussed these methods of analysis in terms of variances of inbred parents, crossbred offspring and the covariance between parents and off-

1. Contribution No. 1304 from Taiwan Agricultural Research Institute.
2. Senior Statistician, Department of Agronomy, TARI, Wufeng, Taichung Hsien, Taiwan 41301, ROC.

spring. Griffing (1956) has classified the four different diallel crossing methods from which at least a p(p-1)/2 set of crosses can be obtained. He also discussed the analysis of each in terms of the variance of genral and specific combining ability under two alternative assumptions with regard to the sampling nature of the experimental material. Theoretically, the important difference among these four methods depends upon whether or not sefings are included. Griffing (1956) suggested that the term "diallel" should be used for those methods which include the sefings and the term "modified diallel" be used for those methods which do not. It was shown that with the modified diallel methods, unbiased estimates of the population variances can be obtained from mean squares of the appropriate analysis of variance (Griffing, 1958). Kempthorne (1956) derived the expectation of mean squares for a complete homozygous diallel. He showed that these mean squares alone could not yield unbiased estimates of the desired parameters. Griffing (1958) also reached the similar conclusion for other diallel methods.

As far as the random model (Eisenhart's Model II) is concerned, the five genetic components of variation defined by Hayman (1954a) are a set of linear transformations of the variances of general and specific combining ability and three other parameters. The combining ability analysis of Griffing (1956, 1958) can be extended to cover the analysis of genetic components of Hayman (1960) by constructing seperate models for selfings and crossbreds for both the homozygous diallel as well as the heterozygous diallel.

Problem of parental heterozygosity in a diallel crosses has scarcely been touched. The intent of this study is to establish basic models that can be further developed to evaluate the complete set of variance components and the genetic constitution of the sampled population.

Models and Analyses

Genotypic Models

The hypotheses listed below either hold in many inheritance mechanisms or are useful simplifications. We assume :

(a) Diploid segregation,

(b) No epistacy,

(c) No multiple allelism.

Let the genotypic values of three genotypes TT, Tt and tt at the t^{th} locus be denoted by d_t, f_t and $-d_t$ respectivly. And let these three genotypes be denoted by a variable θ_t which takes the value 1, 0, and -1 respectively (Hayman, 1954b). Then, if the genotype controlling a metric character is denoted by $\theta=(\theta_1, \theta_2, \cdots, \theta_q)$ and there is no epistacy, the genotypic value is

$$\sum_{t}^{q}[d_t\theta_t+f_t(1-\theta_t^2)] \quad \cdots\cdots (1)$$

as may be seen merely by substituting $\theta_t=1$, 0 and -1 in turn.

Some general results on this representation are ;

(a) The individual with genotype θ_t produces gametes L and t with frequencies $(1+\theta_t)/2$ and $(1-\theta_t)/2$, respectively.

(b) The cross $\theta_i \times \theta_j$ produces progeny TT, Tt and tt with frequencies

$(1+\theta_{it})(1+\theta_{jt})/4$, $(1-\theta_{it}\theta_{jt})/2$ and $(1-\theta_{it})(1-\theta_{jt})/4$,

respectively. These are the probabilities that $\theta_t = 1$, 0 or -1 in the progeny.

(c) In the progency of the cross $\theta_i \times \theta_j$,

$$\begin{aligned}E(\theta_t) &= (1)(1+\theta_{it})(1+\theta_{jt})/4+(0)(1-\theta_{it}\theta_{jt})/2 \\ &\quad +(-1)(1-\theta_{it})(1-\theta_{jt})/4 \\ &= (\theta_{it}+\theta_{jt})/2,\end{aligned}$$

$$\begin{aligned}E(1-\theta_t^2) &= 1-E(\theta_t^2) \\ &= 1-[(1)^2(1+\theta_{it})(1+\theta_{jt})/4+(0)^2(1-\theta_{it}\theta_{jt})/2 \\ &\quad +(-1)^2(1-\theta_{it})(1-\theta_{jt})/4] \\ &= (1-\theta_{it}\theta_{jt})/2\end{aligned}$$

The expected genotypic value of the progeny is

$$E\sum_t^q [d_t\theta_t + f_t(1-\theta_t^2)] = [d_t E(\theta_t) + f_t E(1-\theta_t^2)]$$

$$= \sum_t^q [d_t(\theta_{it}+\theta_{jt}) + f_t(1-\theta_{it}\theta_{jt}]/2$$

(d) The expected genotypic values for selfing and crossbred in the diallel crosses are :

$$\left.\begin{aligned}Y_{ii} &= \sum_t^q [2d_t\theta_{it} + f_t(1-\theta_{it}^2]/2 \\ Y_{ij} &= \sum_t^q [d_t(\theta_{it}+\theta_{jt}) + f_t(1-\theta_{it}\theta_{jt})]/2, \quad (1 \neq j)\end{aligned}\right\} \cdots\cdots (2)$$

In the homozygous diallel the selfings are identical with the parents. This is not so, however, in the heterozygous diallel. The heterozygous×heterozygous selfings show a fall in expected genotypic value equal to $\sum_t^q f_t/2$, as can be proved easily with a one-gene, two-allele model. Thus the selfings in the heterozygous diallel may be regarded as fictitious parents.

Let the frequencies of two allels L and t at the t^{th} locus be deoted by u_t and $v_t (u_t + v_t = 1)$, respectively.

Hence :

$$\left.\begin{aligned}E(\theta_{it}) &= u_t - v_t \\ E(\theta_{it}^2) &= u_t^2 + v_t^2 \\ E(\theta_{it}\theta_{jt}) &= (u_t - v_t)^2 \\ V(\theta_{it}) &= 2u_t v_t \\ V(\theta_{it}^2) &= 2u_t v_t (u_t^2 + v_t^2) \\ V(\theta_{it}\theta_{jt}) &= 4u_t v_t (u_t^2 + v_t^2) - 4u_t^2 v_t^2 \\ Cov(\theta_{it}, \theta_{jt}) &= 0\end{aligned}\right.$$

$$\begin{aligned}
&Cov(\theta_{it}, \theta_{it}^2)=2u_tv_t(u_t-v_t)\\
&Cov(\theta_{it}, \theta_{it}\theta_{jt})=2u_tv_t(u_t-v_t)\\
&Cov(\theta_{it}, \theta_{i't}\theta_{jt})=0\\
&Cov(\theta_{it}, \theta_{jt}^2)=0\\
&Cov(\theta_{it}^2, \theta_{jt}^2)=0\\
&Cov(\theta_{it}^2, \theta_{it}\theta_{jt})=2u_tv_t(u_t-v_t)^2\\
&Cov(\theta_{it}\theta_{jt}, \theta_{it}\theta_{jt})=2u_tv_t(u_t-v_t)^2\\
&Cov(\theta_{it}\theta_{jt}, \theta_{i't}\theta_{j't})=0
\end{aligned} \quad \cdots\cdots (3)$$

By applying the expectations in (3), we have :

$$\begin{aligned}
&\mu_0=E(Y_{ii})=\sum_t^q[d_t(u_t-v_t)+f_tu_tv_t]\\
&\mu_1=E(Y_{ij})=\sum_t^q[d_t(u_t-v_t)+2f_tu_tv_t]\\
&g_i=E(Y_{it})-\mu_1=\sum_t^q[d_t-f_t(u_t-v_t)][\theta_{it}-(u_t-v_t)]/2\\
&z_{ii}=Y_{ii}-\mu_0-2g_i\\
&\quad=\sum_t^q[2f_t(u_t-v_t)\theta_{it}+f_t\theta_{it}^2-f_t+6f_tu_tv_t]/2\\
&s_{ij}=Y_{it}-\mu_1-g_i-g_j\\
&\quad=\sum_t^q[f_t[u_t-v_t)\theta_{it}+f_t(u_t-v_t)\theta_{jt}-f_t\theta_{it}\theta_{jt}-f_t\\
&\qquad+4f_tu_tv_t]/2
\end{aligned} \quad \cdots\cdots (4)$$

The genotypic models in (2) can therefore be vritten as :

$$\begin{aligned}
&Y_{ij}=\mu_0+2g_i+z_{ii}\\
&Y_{ij}=\mu_1+g_i+g_j+s_{ij},\quad (i\#j)
\end{aligned} \quad \cdots\cdots (5)$$

Where :

μ_0 = population mean of the selfings.

μ_1 = population mean of the crossbreds.

$g_i(g_j)$ = general combining ability effect for the i^{th} (j^{th}) parent.

z_{ij} = remaining discrepancy in the i^{th} selfing.

s_{ij} = specific combining ability effect for the cross between the i^{th} and j^{th} parents such that $s_{ij}=s_{ji}$.

It should be pointed out that, by constructing seperate models for selfings and crossbreds, g_i is general combining ability as defined by Sprague and Tatum (1942). Also the s_{ij} fits their definition of specific combining ability, for they defined these effects as deviations from the mean of the F_1's.

Define five genetic components of variation similar to those used by Hayman[1] (1954a) as

$$\left.\begin{aligned} D &= 2\sum_{t}^{q} u_t v_t d_t^2 \\ H_1 &= 2\sum_{t}^{q} u_t v_t (u_t^2 + v_t^2) f_t^2 \\ H_2 &= 4\sum_{t}^{q} u_t^2 v_t^2 f_t^2 \\ F &= 4\sum_{t}^{q} u_t v_t (u_t - v_t) d_t f_t \\ h &= \sum_{t}^{q} u_t v_t f_t \end{aligned}\right\} \quad \cdots\cdots (6)$$

From (3), (4) and (6), it can be shown, after some afgebraic simplifications, that :

$$\left.\begin{aligned} &E(g_i) = E(z_{ii}) = E(s_{ij}) = 0 \\ &\mu_1 - \mu_0 = h \\ &(\mu_1 - \mu_0)^2 = h^2 \\ &V(g_i) = \sigma_g^2 = (D + H_1 - H_2 - F)/4 \\ &V(z_{ii}) = \sigma_z^2 = H_1/4 \\ &Cov(g_i, z_{ii}) = \sigma_{gz} = -(H_1 - H_2)/4 + F/8 \\ &V(s_{ij}) = \sigma_z^2 = H_2/4 \end{aligned}\right\} \quad \cdots\cdots (7)$$

All other covariances are zero.

It is obvious that h, D, H_1, H_2, and F are a set of linear transformations of $\mu_1 - \mu_0$, σ_g^2, σ_z^2, ρ_{gz} and σ_s^2 as mentioned before.

It is seen from (1) that the genotypic value of the s^{th} individual in a random mating population can be expressed by

$$Y_s = \sum_{t}^{q} [d_t \theta_{st} + f_t (1 - \theta^2_{st})].$$

Let the average effect of the t^{th} gene substitution be denoted by

$$\alpha_t = d_t - f_t (u_t - v_t), \quad \text{(Falconer, 1981)}$$

then the breeding value and the dominance deviation of the s^{th} individual can be expressed by

$$\begin{aligned} b_s &= \sum_{t}^{q} [(\theta_{st} - 1) u_t \alpha_t + (\theta_{st} + 1) v_t \alpha_t] \\ &= \sum_{t}^{q} \alpha_t [\theta_{st} - (u_t - v_t)] \end{aligned}$$

---

[1]The five components (h, D, H_1, H_2 and F) defined here are not same as those (h', D', H_1', H_2' and F') defined by Hayman (1954a). The relationship between these two sets of components is :

$h' = 4h$, $D' = 2D$, $H_1' = 2(H_1 + H_2)$, $H_2' = 4H_2$, $F' = 2F$.

and

$$d_s=[Y_s-E(Y_s)]-b_s$$
$$=\sum_t^q f_t[2u_tv_t+(u_t-v_t)\theta_{st}-\theta^2_{st}]$$

respectively.

By applying (3) and (6), we have

$$\left.\begin{aligned} V(b_s)&=\sigma_a^2=D+H_1-H_2-F\\ V(d_s)&=\sigma_d^2=H_2\\ Cov(b_s,\ d_s)&=0\end{aligned}\right\} \quad \cdots\cdots (8)$$

Where σ_a^2 and σ_d^2 are the additive and dominance variances respectively. From (7) and (8), we have derived the relations that:

$$\left.\begin{aligned} \sigma_a^2&=4\sigma_g^2\\ \sigma_d^2&=4\sigma_s^2\end{aligned}\right\} \quad \cdots\cdots (9)$$

This result (9) is to be distinguished from that of a homozygous diallel. In a homozygous diallel $\sigma_a^2=2\sigma_g^2$ and $\sigma_d^2=\sigma_s^2$.

Genetic Constitution of Sampled Population

The genetic components of variation can be estimated from the proposed models. Certain properties of the population under investigation can then be investigated in the following manner.

(a) Mean level of dominance

Mean level of dominance can be estimated by the expression

$$\sqrt{(H_1+H_2)/D.}$$

This reduces to

$$\sqrt{2\Sigma u_tv_tf_t^2/2\Sigma u_tv_td_t^2}$$

which becomes $\bar{f}/\bar{d}$ if the q loci have equal effects. Whereas if the loci have unequal effects, the estimated dominace ratio is automatically weighted in favor of those with largest effects and those with allelic frequency nearset to 0.5. As unequal effects may be the usual case and we can only deal with the combined effects of all the loci, the loci with greater effects will contribute more information to the weighted estimate than the loci at low frequency and/or with small effects (Dickinson and Jinks, 1956).

In addition to the above, with similar reasoning

$$\sqrt{H_2/D^2}=\sqrt{4\Sigma u_t^2v_t^2f_t^2/4\Sigma u_t^2v_t^2d_t^4}=\bar{f}/\bar{d}^2$$

These estimates of $\bar{f}/\bar{d}$ and $\bar{f}/\bar{d}^2$ can be solved simultaneously for $\bar{f}$ and $\bar{d}$.

(b) Average level of heterozygosity and allelic frequencies The expression

$$H_2/(H_1+H_2)$$

gives a weighted estimate of the average level of heterozygosity. This reduced to

$$4\Sigma u_t^2 v_t^2 f_t^2 / 2\Sigma u_t v_t f_t^2$$

which becomes $2\bar{u}\bar{v}$ if the q loci have equal effects. Whereas if the loci have unequal effects, this estimate will be the average level of the heterozygosity at loci showing dominance since the ratio is confined to terms in f^2, The loci showing greater dominance will also contribute relatively more to this weighted estimate than those showing little dominance. In addition, this estimate will be weighted in favor of loci with allelic frequency nearest to 0.5. Hence biased estimation of the average heterozygosity will occur whenever the level of heterozygosity differes from one locus to another which may in reality be the situation. In addition to the above. $\bar{u}+\bar{v}=1$. These equations of $\bar{u}$ and $\bar{v}$ can be solved simultaneously for $\bar{u}$ and $\bar{v}$.

(c) Number of loci showing diminance

The expression

$$4h^2/H_2 = 4(\Sigma u_t v_t f_t)^2 / 4\Sigma u_t^2 v_t^2 f_t^2$$

gives an estimate of the number of loci showing dominance provided that certain assumptions are made. Assuming that q′ loci involved exhibit unidirectional dominance and have equal effects, the expression becomes

$$4(\Sigma u_t v_t f_t)^2 / 4\Sigma u_t^2 v_t^2 f_t^2 = (q' u_t u_t f_t)^2 / q' u_t^2 v_t^2 f_t^2 = q'^2/q' = q'$$

An underestimate of q′ will occur if either or both assumptions are not valid. Thus, when all loci do not exhibit unidirectional dominance, positive and negative values of f cancel out in h, with a consequent reduction in the estimated value of q′. In addition, inequality of effects at the q′ loci will also give a lower estimate. Since the present estimate is confined to a ratio of terms in f, it obviously gives a lower bound of the number of loci showing doming dominande and is biased in favor of those of greater dominance (Dickinson and Jinks 1956).

(d) Dominance level over loci

An estimate of this is given by the expression

$$F/2\sqrt{D(H_1 - H_2)}$$

provided that certain assumptions are made. If the ratio of f to d and the difference of u and v are constants over all loci, then the absolute value of this expression will be 1. However, if either or both assumptions are not valid, then the absolute value of this expression will be close to 0. Hence, regardless of the dominance level this ratio can be used to measure the extent to which the dominance level varies from one locus to another (Mather and Jinks, 1971).

Conclusion

Development of an effective breeding program for a given species depends upon knowledge of the kinds and magnitudes of variability which contribute to differences among individuals. Additive, dominance and epistatic genetic variances are particularly important in understanding the inheritance mechanisms of the economically important

characters.These genetic variances can only be estimated from covariance between relatives (e. g., between parent and offspring, between full-sibs, between half-sibs, etc.). Covariances between relatives can only be estimated from experiments based on controlled crosses with specific mating designs. Due to restrictions of time, space, cost and biological limitations, some of the mating designs for estimating genetic variances are confined to those dealing with one common generation of parental materials originating as a random sample from a single random population. Nested, factorial and diallel designs are well known and are commonly used. All of the above designs permit estimates of covariance between full-sibs and between half-sibs which can be interpreted in terms of genetic variances of the sampled population. Although diallel designs are relatively difficult to use in some breeding program, they should be particularly useful when a complete evaluation of variance components is desired including those due to selfing, maternal and reciprocal effects.

Genetic variances are estimated from the covariances between relatives which in turn are estimated from the functions of variance components. Therefore efficient estimates of genetic variances depends entirely on efficient estimates of variance components.

Constructing seperate models for selfings and crossbreds besed on their genotypic values, the genetic components analysis of Hayman and the combining ability analysis of Griffing can be combined to provide estimates of various genetic parameters of the sampled population from a heterozyous diallel.

In practice, estimates of the genetic components of variation and other parameters derived from them may not be very precise. This is not only because of the lack of knowledge about the true genetic mechanisms that control the inheritance of the traits studied but also because of the sizable amount of of environmental variation usually associated with the measurements themselves. Interpretations and applications of these genetic parameters should be carefully made. Different traits may need different models to explain their variation and incompatible results may indicate the incorrectness of the model used.

Literatura Cited

Dickinson, A. G. and J. L. Jinks. 1956. A generalized analysis of diallel crosses. Genetics 41 : 65-78.

Falconer, D. S. 1981. Introduction to quantitative genetics. 2nd Ed., The Ronald Press, New York. 340pp.

Griffing, B. 1956. Concept of general and specific combining ability in relation to diallel crossing systems. Aust. J. Biol. Sci. 9 : 463-493.

Griffing, B. 1958. Application of some sampling variables in the identification of methods which yield unbiased estimates of genotypic variance components. Aust. J. Biol. Sci. 11 : 219-245.

Hayman, B. I. 1954a. The analysis of variance of diallel tables. Biometrics 10 : 234-244.

Hayman, B. I. 1954b. The theory and analysis of diallel crosses. Genetics 39 : 789-809.

Hayman, B. I. 1958. The theory and analysis of diallel crosses. II. Genetics 43 : 63-85.

Hayman, B. I. 1960. The theory and analysis of diallel crosses. III. Genetics 45 : 155-172.

Jink, J. L. 1954. The analysis of heritable variation in a diallel cross of *Nicotiana rustica* varieties. Genetics. 39 : 767-788.

Jinks, J. L. and B. I. Hayman. 1953. The analysis of diallel crosses. Maize Genetics News Letter 27 : 48-54.

Kempthorne, O. 1956. The theory of the diallel cross. Genetics 41 : 451-459.

Mather, K. and J. L. Jinks. 1971. Biometrical genetics. 2nd Ed., Cornell Univ. Press, New York. 382pp.

Sprague, G. F. and L. A. Tatum. 1942. General versus specific combining ability in single crosses of corn. J. Amer. Soc. Agron. 34 : 923-932.

Yates, F. 1947. Analysis of data from all possible reciprocal crosses between a set of parental lines. Heredity 1 : 287-301.

異質型全互交之成分分析[1]

劉　　清[2]

摘　　要

本文依據異質型全互交自交後代與雜交後代不同基因型值分別建立其數學模式以證明在全互交分析上 Hayman 氏之遺傳成分分析法與 Griffing 氏之組合力分析法基本上爲等義的，且綜合此二種分析法來分析異質型全互交以對取樣族羣提供各項遺傳介量之估值。文中亦指出異質型全互交與同質型全互交在遺傳介量估算上之差異。

1.臺灣省農業試驗所　研究報告第 1304 號。

2.本所農藝系研究員。臺灣省　臺中縣　霧峰鄉。